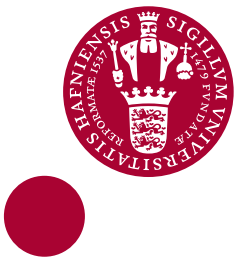

The Weaponization of Social Media

Spear Phishing and Cyberattacks on Democracy

Bossetta, Michael

Published in:
Journal of International Affairs

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Bossetta, M. (2018). The Weaponization of Social Media: Spear Phishing and Cyberattacks on Democracy. *Journal of International Affairs*, *2018 Special Issue, vol. 71*(2), 97-106. [6].

THE WEAPONIZATION OF SOCIAL MEDIA: SPEAR PHISHING AND CYBERATTACKS ON DEMOCRACY

Michael Bossetta

Abstract: State-sponsored cyber groups have long utilized spear phishing to pierce government networks. Spear phishing relies on social engineering to trick individuals into revealing sensitive information or downloading malicious software, rather than hacking into a system vulnerability by force. While email remains the preferred medium to conduct spear-phishing attacks, social media has opened up new attack vectors for politically motivated cyberattacks. Social media platforms, as high-trust environments typically accessed from a mobile device for personal entertainment or networking, are highly conducive waters for spear phishing. Moreover, the wealth of public information available on social media can be exploited by threat actors to devise sophisticated (and automated) spear phishing campaigns that target government and military personnel. This study examines how illiberal regimes are weaponizing social media to conduct spear phishing and cyber espionage against Western governments. A theoretical model of spear phishing on social media is proposed and supported by recent empirical examples from the European Union and United States.

Much of the controversy around Russian interference in the 2016 U.S. election has focused on state-sponsored attempts to manipulate public opinion through social media. However, just weeks after President Donald J. Trump's inauguration, Russian operatives demonstrated a cyber capability far exceeding the paid use of "trolls" to spread propaganda. More than 10,000 tweets—each laced with hyperlinks containing malware—were sent directly to U.S. Defense Department employees on Twitter.[1] The messages were tailored to appeal to the employees' individual interests and generated click rates nearing 70 percent. In some cases, employees' family members were targeted, and devices containing sensitive government information were compromised through shared home Wi-Fi networks. ZeroFOX, a leading cybersecurity firm, referred to the malicious micro-

targeting campaign as "the most well organized, coordinated attack at the nation-state level we've ever seen…it's a harbinger of things to come."[2]

Increasingly, foreign actors are turning to social media to carry out cyberattacks. With an estimated 3.2 billion people active on social media, state-affiliated threat groups have access to massive troves of personal data that can inform sophisticated spear phishing campaigns.[3] Moreover, social media platforms open up new attack vectors, and advances in technology displace the notion that social media is difficult to weaponize for country-specific policy goals.[4] This paper outlines how illiberal regimes are weaponizing social media to carry out cyberattacks against Western governments and their personnel. A theoretical model of social media spear phishing is proposed and supported empirically with recent examples from Great Britain, France, Germany, and the United States.

Spear Phishing, Cyber Espionage, and Social Media

Spear phishing is a targeted phishing attack customized to an individual or set of individuals. Phishing attacks bait victims to take an action, which typically involves clicking a malicious link or opening an email attachment that harbors a malware payload.[5] Both actions can lead victims to fabricated websites that ask for login credentials ("credential spear phishing")[6] or download software directly to the victim's device ("drive-by downloads").[7] Attackers then leverage the credentials or infected devices to gain access to a broader network, stealing information and often remaining undetected for extended periods of time.

Cyberespionage by state-affiliated actors accounts for 25 percent of successful phishing breaches.[8] The primary motivation is to extract sensitive government data, which can be appropriated for several pernicious purposes. As shown by the Russian-backed spear phishing attacks on the email accounts of Hillary Clinton's campaign manager, John Podesta, and former U.S. Secretary of State, Colin Powell,[9] the data can be publicly released for defamation ("doxxing") or electoral influence. The theft of intellectual property, such as military plans or technology innovations, can be utilized to advance strategic geopolitical objectives. Moreover, the procurement of trade or manufacturing information can provide a tactical advantage in trade negotiations.

Although email remains the preferred attack vector for spear phishing, spear phishing attacks on social media increased 500 percent in 2016.[10] The attacks peaked around major events like the Olympics and the U.S. election, as threat actors aimed to exploit public interest in trending online conversations. As with email, the majority of social media phishing attacks are financially motivated cybercrimes. However, government agencies are becoming increasingly aware of the political and military risks associated with social media weaponization.[11] In

the following sections, I develop a theoretical model of spear phishing on social media to highlight how illiberal regimes are weaponizing social media platforms to attack Western democracies.

A Model of Spear Phishing on Social Media

The model consists of five phases: Collect, Construct, Contact, Compromise, and Contagion. The first phase of a spear phishing attack is to collect data on the intended target. Social media platforms offer a wealth of publicly available data, and these data can be exploited to then construct fake accounts that appeal to the target's personal or professional interests. Using these accounts, attackers contact targets through any variety of communicative modes enabled by the platform, ranging from friend requests to direct messages to targeted advertisement campaigns. Depending on the attacker's intentions, the target may be tricked into revealing information or clicking a link that compromises the target's account or device. If successful, the attack can then induce a contagion effect, magnifying its scope and putting others at risk. Below, I outline each of these phases in detail.

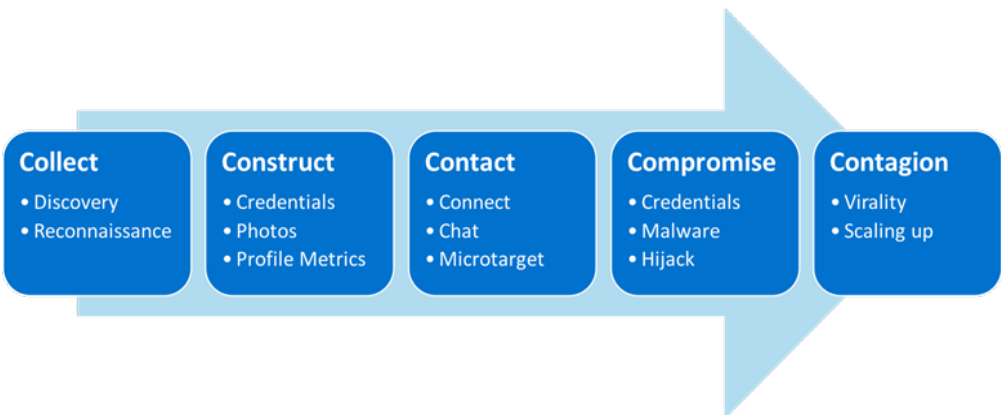

Figure 1: Model of Spear Phishing on Social Media

Collect

To increase the likelihood of success for a spear phishing attack, threat actors first collect data to inform their operations.[12] This data collection can be divided into two sub-categories: discovery and reconnaissance. Discovery entails the use of social media data to identify targetable persons. Threat actors may know the agency they wish to breach, but they may not have a precise list of individuals to target. Simply by using a platform's search function, specific keywords relating to an organization or policy area can be queried to discover individuals who may possess knowledge or credentials deemed valuable. In 2015, the British secu-

rity agency, MI5, warned employees that Chinese and Russian spies were using LinkedIn, the professional networking platform, to identify government employees to recruit for espionage operations.[13]

Once identified, reconnaissance can be carried out on the target. In addition to identifying one's online connections, threat actors can collect personally identifiable information such as email addresses, phone numbers, work history, education, or interests. The attacker can also observe the target's previous online interactions, particularly on open platforms such as Twitter and LinkedIn. Ahead of the 2017 French election, Facebook identified approximately two dozen fake accounts spying on then-candidate Emmanuel Macron's presidential campaign. These accounts, linked to the Russian hacking group Fancy Bear—the same organization responsible for the email hacks of the U.S. Democratic National Committee—were "posing as friends of friends of Macron associates and trying to glean information from them."[14] Although Facebook detected the operation early and blocked the accounts, the apparent intent of the campaign was to gather intelligence for a spear-phishing attack, "…to get targets to download malicious software or give away login information."[15]

The purpose of data collection is to identify and observe targets to design a customized attack, thereby increasing its chances for success. This data collection process has both manual and automated variants. The former is time intensive and likely reserved for high-profile targets (i.e. "whaling"), but data can also be automatically collected at scale through platforms' Application Programming Interfaces (APIs) or commercial third-party software such as Grouply, which harvests personal information from public Facebook groups.[16] Additionally, programs can be written in open-source programming software, like Python or R, to compile employee lists from agencies' LinkedIn accounts or web pages.[17,18] From there, employees' personal information can be harvested automatically across several different social media accounts.

Construct

Once threat actors collect data, they construct fake social media profiles to interact with the target. The design of these profiles is informed by the data collection and seeks to establish common ground. The constructed persona may include fabricated credentials, such as working in the same organization or having attended the same university. Constructed accounts may even mimic or fabricate organizations, as demonstrated by the Russian Internet Research Agency's creation of Facebook pages such as "United Muslims of America" and "Blacktivist."[19]

Given that defense contractors, military service members, and IT-personnel typically skew male, constructed accounts are often portrayed as attractive females,

whose photos are stolen from authentic profiles ("Catphishing"). This method formed the basis of the Robin Sage Experiment, where a cybersecurity firm leveraged the constructed profile of a "young, attractive, and edgy female" to establish LinkedIn connections with senior executives at the U.S. National Security Agency, Department of Defense, and military intelligence groups.[20] Iranian hackers, targeting Middle Eastern industries and governments, put the Robin Sage experiment into practice by creating the fictitious "Mia Ash."[21] To bolster the persona's authenticity, the group repurposed the photos of a female Romanian photographer and established accounts on LinkedIn, Facebook, Twitter, and Instagram.

Recently, the German domestic intelligence service Bundesamt für Verfassungsschutz (BfV) released screenshots of fake LinkedIn accounts constructed by Chinese intelligence services.[22]

The examples illustrate the accounts' clear orientation toward Western users working with issues related to Chinese foreign policy.

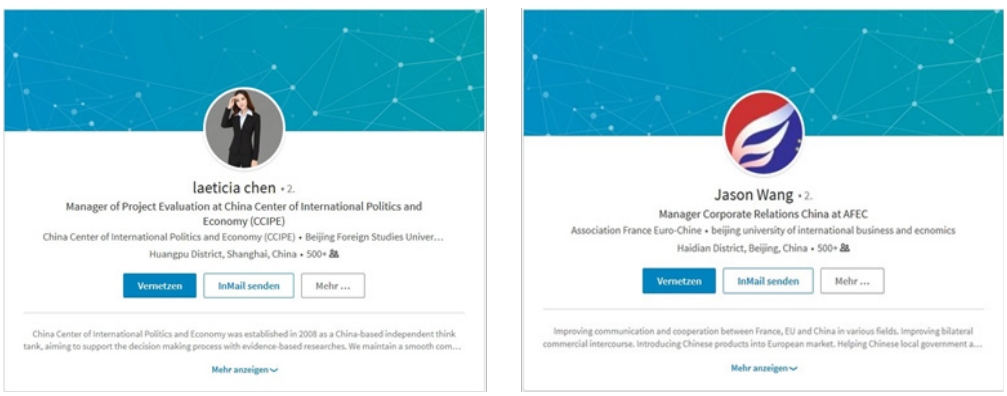

Figure 2: Constructed LinkedIn Profiles (Source: BfV)

Perhaps even more crucial than credentials and photos for a constructed account are profile metrics: the platform-specific, quantitative indicators that users rely on to make judgements about a profile. Chief among these indicators is the number of connections associated with an account; these "friend cues" signal endorsements of authenticity from other users on the platform. State-sponsored actors recognize the importance of profile metrics, and they have developed innovative methods to artificially inflate them.[23] On LinkedIn, for example, an Iranian group utilizes networks of "Leader" and "Support" accounts.[24] The purpose of Support accounts is to connect with, and leave public endorsements for, the Leader accounts (who maintain more than 500 connections, which is the maximum number publicly displayed on LinkedIn's interface). Leader accounts may alter their front-facing identities by changing their names and photos periodically, but they still retain

their existing connections and networks. Another example of profile metrics is the account creation date, which is featured prominently on Twitter. Older accounts are more likely to be considered authentic, and therefore fabricated accounts, particularly those of bots, can lie dormant for years—"aging"—before being activated for malicious purposes.[25]

Contact

With data-driven, constructed accounts created, threat actors seek to initiate contact with their targets. This can be done in several ways contingent upon the platform's "digital architecture," defined as "the technical protocols that enable, constrain, and shape user behavior in a virtual space."[26] One common method is to connect with the target's account: "friending" on Facebook, "connecting" on LinkedIn, or "following" on Twitter and Instagram. Pending the default privacy settings of the platform and the users' customization of them, an accepted request can reveal non-public data on the target. Another contact method is the platform's specific chat services, such as Facebook Messenger, LinkedIn InMail, or Twitter Direct Message. Through these chat services, threat actors can request information or send malicious links in an environment where targets are accustomed to engaging in discussions with close friends.

A third, higher-risk contact method is advertisement campaigns, where threat actors can pay platforms to target certain demographics of users, such as those in a geographic location, holding a particular position, or working for a designated organization. Although social media companies hire human moderators to manually approve paid advertisements, there are documented instances of malware links bypassing this vetting process.[27] With advances in technology and the sophistication of threat groups, however, malicious content can even be microtargeted without the use of advertisement services. As alluded to in the introduction, Russian operatives utilized Twitter to target U.S. Defense Department employees with individualized messages containing malicious links.

A recent project by ZeroFOX researchers sheds insight into how such an attack could be conducted.[28] First, a list of Twitter accounts is seeded into a program, which then automatically collects recently issued tweets from each account. Then, machine learning algorithms generate custom tailored messages—based on the content of the harvested tweets—and send them directly to the target. The message includes a malicious link shortened with Google's popular URL shortener, which disguises the link's destination, increases trustworthiness through a well-known brand, and allows the sender to monitor if the link has been clicked. The malicious tweets can then be sent at scheduled, strategic time points when the target is predetermined to be most active on Twitter. This process can be

entirely automated and was able to generate click through rates (CTR) as high as 66 percent, meaning that two-thirds of the generated links were clicked. The Russian attack, likely using a similar method, targeted more than 10,000 Defense Department employees and yielded 7,000 clicks for a CTR of approximately 70 percent.[29] To put that in perspective, the median CTR for phishing emails typically ranges between 6-13 percent, depending on the industry.[30]

Compromise

Compromise refers to the outcome of a successful spear phishing attack. 95 percent of phishing breaches install software on the target's device, and this is typically achieved when a user downloads an email attachment that contains a malware payload.[31] On social media, where file attachments are uncommon, attackers aim to redirect users to a URL that installs malware on the device when accessed. Once the malware is installed, threat actors use the compromised device as a beachhead, crawling through a network and stealing information. Often, the target is unaware that their device has been affected, and attackers can remain undetected in a network for years.[32]

However, some threat actors prefer to pivot off the attack quickly by hijacking the social media account. The ISIS-affiliated "CyberCaliphate," for example, hacked U.S. Central Command's Twitter and YouTube accounts in 2015, using the accounts to briefly spew propaganda.[33] More sophisticated attacks, such as the Russian one outlined above, can potentially compromise thousands of accounts at once, simultaneously reporting disinformation that could disrupt the news cycle or wreak havoc on the stock market. Other motivations for compromising social media accounts can be to steal private messages for blackmail or doxxing. Moreover, state-sponsored actors can generate messages from these accounts to stage evidence against individuals. Iran's Islamic Revolutionary Guard Corps (IRGC) has targeted the Facebook accounts of dissidents and journalists, shortly before arresting them for espionage. The IRGC sent messages from these accounts to other targets, presumably to entrap them by using the interactions as court evidence for participation in an American-affiliated spy ring.[34]

Contagion

The malicious use of compromised accounts can lead to viral contagion and magnify the scope of an attack.[35] Threat actors can launch attacks directly from compromised accounts, targeting the victim's connections through private messages. Turkish threat actors employed this technique by first compromising the Twitter account of the Indian ambassador to the United Nations, then the president of the World Economic Forum, and eventually several high-profile U.S.

journalists.[36] Private messages between Fox News journalists and President Trump were among the stolen data. This means that, in theory, the President of the United States' Twitter account was susceptible to compromise.

Contagion is especially dangerous because threat actors can target vulnerable victims and scale up to bigger targets. The Iranian hacking of the U.S. State Department in 2015, for example, used the compromised Facebook accounts of young government employees to infect others higher up in the administration.[37] Furthermore, contagion can be facilitated by algorithmic recommender systems that promote compromised accounts to others.[38] Finally, it is worth noting that contagion poses a particularly acute threat as technology moves toward the Internet of Things, where networks are increasingly interconnected and do not require user action to connect with one another.

Conclusion

Many cybersecurity reports do not explicitly discuss social media as an attack vector to breach government networks. Therefore, a lack of understanding remains regarding the role of social media platforms in politically motivated cyberattacks. This paper has shed light on how state-affiliated threat actors weaponize social media platforms to execute spear phishing campaigns. Using sophisticated social engineering tactics, maliciously aligned actors seek to manipulate individuals through seemingly innocuous interactions on social media. Democracy is under attack when its fundamental tenant—trust—is exploited for such illiberal pursuits.♔

Michael Bossetta *is a PhD candidate in the Department of Political Science at the University of Copenhagen. His research uses computational methods to investigate politicians' and citizens' use of social media during elections. He is the producer and host of the Social Media and Politics Podcast, available on Apple Podcasts, Spotify, and all downloadable podcasting apps. You can follow him on Twitter @MichaelBossetta and the podcast @SMandPPodcast.*

NOTES

1 Massimo Calabresi, "Inside Russia's Social Media War on America," *Time*, 29 May 2017.

2 ZeroFOX Research, "Russia just used Trump's Favorite Social Network to Hack the U.S. Government," *ZeroFOX Social Media Security Blog*, 18 May 2017, https://www.zerofox.com/blog/russia-just-used-trumps-favorite-social-network-hack-us-government/.

3 Simon Kemp, "Digital in 2018: World's Internet Users Pass the 4 Billion Mark," *We Are Social (Blog)*, 30 January 2018, https://wearesocial.com/blog/2018/01/global-digital-report-2018.

4 Clay Shirky, "The Political Power of Social Media: Technology, the Public Sphere, and Change," *Foreign Affairs* 90, no. 1 (2011), 31.

[5] Indranil Bose and Alvin Chung Man Leung, "Unveiling the Mask of Phishing: Threats, Preventive Measures, and Responsibilities," *Communications of the Association for Information Systems* 19, no. 1 (2007), 525.

[6] Grant Ho et al., "Detecting Credential Spearphishing Attacks in Enterprise Settings," (Vancouver, Canada: Proceedings of the 26th USENIX Security Symposium, 16-18 August 2017), 469.

[7] Amir Javed, Pete Burnap, and Omer Rana, "Prediction of Drive-By Download Attacks on Twitter," *Information Processing and Management*, https://doi.org/10.1016/j.ipm.2018.02.003.

[8] Verizon, 2017 Data Breach Investigations Report: 10th Edition, http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/.

[9] Lorenzo Franceschi-Bicchierai, "How Hackers Broke into John Podesta and Colin Powell's Gmail Accounts," *Motherboard*, 20 October 2016, https://motherboard.vice.com/en_us/article/mg7xjb/how-hackers-broke-into-john-podesta-and-colin-powells-gmail-accounts.

[10] Proofpoint, Inc., *Q4 2016 & Year in Review: Threat Summary*, https://www.proofpoint.com/us/threat-insight/threat-reports.

[11] Scott Solomon, "Social Media: The Fastest Growing Vulnerability to the Air Force Mission," *Air Force Research Institute*, January 2017.

[12] Katharina Krombholz et al., "Advanced Social Engineering Attacks," *Journal of Information Security and Applications* 22, (2015), 115.

[13] Ian Drury and David Williams, "Foreign Spies on LinkedIn trying to Recruit Civil Servants by 'Befriending' them before Stealing British Secrets," *Daily Mail*, 9 August 2015.

[14] Joseph Menn, "Exclusive: Russia used Facebook to Try to Spy on Macron Campaign – Sources," *Reuters*, 27 July 2017.

[15] Ibid.

[16] Grouply, http://grouply.io/.

[17] Yannick Scheelen and Daan Wagenaar, "The Devil is in the Details: Social Engineering by Means of Social Media," *Research Project: University of Amsterdam*, 11 July 2011, http://work4.delaat.net/rp/2011-2012/p60/report.pdf.

[18] Matthew Edwards et al., "Panning for Gold: Automatically Analyzing Online Social Engineering Attack Surfaces," *Computers & Security* 69 (2017): 18-34.

[19] United States of America v. Internet Research Agency, LLC, 18 U.S.C. (2018).

[20] Thomas Ryan, "Getting in Bed with Robin Sage," *Provide Security*, 2010, https://media.blackhat.com/bh-us-10/whitepapers/Ryan/BlackHat-USA-2010-Ryan-Getting-In-Bed-With-Robin-Sage-v1.0.pdf.

[21] Counter Threat Unit Research Team, "The Curious Case of Mia Ash: Fake Persona Lures Middle Eastern Targets, *Secureworks*, 27 July 2017, https://www.secureworks.com/research/the-curious-case-of-mia-ash.

[22] Bundesamt für Verfassungsschutz, "Vorsicht bei Kontaktaufnahme über Soziale Netzwerke – Fortschreibung," B*undesamt für Verfassungsschutz*, 12 December 2012, https://www.verfassungsschutz.de/de/aktuelles/sicherheitshinweise/sh-20171212-kontaktaufnahme-ueber-soziale-netzwerke-fortschreibung.

[23] Arun Vishanath, "Getting Phished on Social Media," *Decision Support Systems* 103 (2017), 74.

[24] Dell Secureworks Counter Threat Unit Threat Intelligence, "Hacker Group Creates Network of Fake LinkedIn Profiles," *Secureworks*, 7 October 2015, https://www.secureworks.com/research/suspected-iran-based-hacker-group-creates-network-of-fake-linkedin-profiles.

[25] ZeroFOX Research, "Social Engineering in the Social Media Age: Top Fraudulent Account & Impersonator Tactics," *ZeroFox Research* (2016), 18.

[26] Michael Bossetta, "The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election," J*ournalism & Mass Communication Quarterly* 90, no. 2 (2018), 3, https://doi.org/10.1177/1077699018763307

[27] ZeroFOX Research, "Social Engineering," 15.

[28] John Seymour and Philip Tully, "Generative Models for Spear Phishing Posts on Social Media"

(Long Beach, USA: *Electronic Proceedings of the 31st Conference on Neural Information Processing Systems*, 4-9 December 2017).

[29] Sheera Frenkel, "Hackers Hide Cyberattacks in Social Media Posts," *New York Times*, 28 May 2017.

[30] Verizon (2017), 11.

[31] Verizon (2017), 32

[32] Verizon (2017), 48.

[33] Dan Lamothe, "U.S. Military Social Media Accounts Apparently Hacked by Islamic State Sympathizers," *Washington Post*, 12 January 2015.

[34] David Sanger and Nicole Perlroth, "Iranian Hackers Attack State Dept. via Social Media Accounts," *New York Times*, 24 November 2015.

[35] Arun Vishwanath, "Diffusion of Deception in Social Media: Social Contagion Effects and its Antecedents," *Information Systems Frontiers* 17, no. 6 (2015), 1353-1367.

[36] ZeroFOX Research, "Social Media Hacking Campaign Targets Journalists through Direct Messages," *ZeroFOX Social Media Security Blog*, 26 January 2018, https://www.zerofox.com/blog/social-media-hacking-campaign-targets-journalists-direct-messages/.

[37] Sanger and Perlroth (2015).

[38] Marco Balduzzi et al., "Abusing Social Networks for Automated User Profiling," In Jha Somesh, Robin Summer, and Christian Kreibich, eds., *Recent Advances in Intrusion Detection*, (Berlin: Springer, 2010).